NAPHEN, CASSANDRA NICHOLE, M.S. Targeted Analogue Detection from *Pestalotiopsis microspora* using Molecular Networking and Mass Defect Filtering: Towards the Development of new Anti-Virulence Leads against Methicillin Resistant *Staphylococcus aureus*. (2019)
Directed by Dr. Nadja B. Cech. 49 pp.

Effective tools are needed to enable prioritization of lead molecules in natural products drug discovery research. As the field of mass spectrometry continues to advance, more people have access to high-resolution instruments, and more researchers rely on this technology for metabolite and compound class identification. To effectively access lead molecules from natural product mixtures, it is important to comprehensively assess large datasets in the most timely and cost-effective manner. For analyses based on liquid chromatography-mass spectrometry (LC-MS), several post-acquisition data analysis approaches can be employed to identify analogues of a lead molecule of interest. With this study, we sought to compare the effectiveness of mass defect filtering and molecular networking for this purpose. Secondary metabolite production by the ambuic acid-producing endophytic fungus *Pestalotiopsis microspora* was used as a test case for these analyses. Eight putative analogues of ambuic acid were identified using mass defect filtering and molecular networking analyses. Our results illustrate advantages and disadvantages of molecular networking and mass defect filtering and suggest that the combination of both data mining techniques may enable the most comprehensive evaluation of analogues.

TARGETED ANALOGUE DETECTION FROM *PESTALOTIOPSIS MICROSPORA* USING MOLECULAR NETWORKING AND MASS DEFECT FILTERING: TOWARDS THE DEVELOPMENT OF NEW ANTI-VIRULENCE LEADS AGAINST METHICILLIN RESISTANT *STAPHYLOCOCCUS AUREUS*

by

Cassandra Nichole Naphen

A Thesis Submitted to
the Faculty of The Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Greensboro
2019

Approved by

_____
Committee Chair

APPROVAL PAGE

This thesis written by CASSANDRA NICHOLE NAPHEN has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair _____

Committee Members _____

_____

_____

_____November 14, 2018_____
Date of Acceptance by Committee

____November 14, 2018_____
Date of Final Oral Examination

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

1.1 Antibiotic Resistance

Over 2 million people are diagnosed with an antibiotic-resistant infections in the United States each year, and these infections are responsible for 23,000 annual fatalities.[1-2] The bacterial pathogens currently considered to be of the highest concern by the CDC are known as the ESKAPE pathogens*: Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa*, and *Enterobacter* species. These pathogens are responsible for many of hospital-acquired infections in the United States and are becoming resistant to many, or even all, of the antibiotic treatments currently available.[1, 3-7]

1.2 Methicillin- Resistant *Staphylococcus aureus* (MRSA)

The most prominent and costly skin and soft tissue infections are caused by the Gram-positive organism *Staphylococcus aureus.*[8-10] Methicillin-resistant *Staphylococcus aureus* (MRSA) possesses an arsenal of virulence factors that are controlled by a density-dependent regulatory system known as quorum sensing.[11-14] Quorum sensing systems allow bacterial cells in close proximity to one another to communicate, activate virulence factors contributing to pathogenesis, and more effectively colonize host tissue.[11, 15-17]

While over-prescription of antibiotics or lack of antibiotic stewardship[2] leads to the development of antibiotic resistance in the target organism, targeting and inhibiting the pathways that produce these virulence factors is a promising strategy for disarming the pathogen. By avoiding the use of antibiotics, this approach could facilitate clearance of the infection while lowering the risk of antibiotic resistance development and preserving the beneficial microflora.[13, 18-21]

1.3 Targeting Anti-virulence for Drug Discovery

MRSA quorum sensing is controlled by the accessory gene regulator (*agr*) system. There are three main protein targets in the *agr* system for small molecule inhibition: AgrC, AgrA, and AgrB. Much research thus far has focused on finding inhibitors that target the transmembrane protein AgrC, but this target varies greatly even among bacterial strains and has shown evidence of being hypermutable.[12, 14, 21-25] The AgrB protein of the *agr* system is a more favorable target because it is highly conserved among Gram-positive pathogens.[22] To date, the fungal metabolite ambuic acid is the only reported *agr* signal biosynthesis inhibitor that functions via AgrB inhibition.[26-29] Ambuic acid is reliably produced by the Ascomycete fungus belonging to order Xylariales, *Pestalotiopsis microspora*.[30-32] Quorum sensing inhibition by ambuic acid was shown *in vivo* to reduce MRSA-induced abscess formation in a mouse model by half with a 5-µg dose.[26] Given the success of these initial studies, there is interest in identifying additional analogues of ambuic acid that might possess higher potency and/or provide insight into the mechanism of action for this compound.

1.4 Data Mining via Mass Defect Filtering and Molecular Networking

There is often a priority for the isolation of analogues to conduct structural activity relationship studies and identify novel chemical entities. To effectively mine large mass spectrometry datasets originating from complex fungal mixtures, data processing techniques capable of prioritizing lead ions are often used.[33-39] Data processing techniques such as molecular networking and mass defect filtering represent commonly used approaches for dereplicating and prioritizing lead molecules that have structural relatedness to the desired compound within large datasets and can easily be incorporated into natural products discovery workflows.[34, 39-41] These tools are valuable for assessing whether target ions are present and can give insight into whether a specific organism should be pursued for natural product discovery.

Mass defect filtering is a simple tool to identify compounds that have similar molecular formulae, and as such, are likely to have similar chemical structures. High resolution mass spectrometry has the confidence in four decimal places after the nominal mass and allows for higher ability to prediction of molecular formula.[42-43] This principles of isotopic abundance that each element contributes to this exact mass and the mass defect. A filter can be placed on the full scan mass spectra data to isolate ions that only fall within an user-defined window around the precursor mass. The absolute value of the mass defect of an ion reflects the ion's elemental composition because each element has a unique finite mass defect. The mass defect of a given molecule is defined as the deviation between its nominal mass and its exact monoisotopic mass. Numerous studies have

utilized mass defect filtering to identify molecules likely to be structurally related to a target analyte or a core substructure in an entire mass spectrometry dataset.[34, 43-46] Mass defect filtering has also been utilized for studies involving drug metabolism,[42-43] isotope labeling experimentation and removal of media or polymer components.[47-48]

Another commonly utilized data mining tool, global natural products social molecular networking (GNPS), is a natural product and metabolomics crowdsourced analysis platform with public reference libraries, public data repositories, and living data.[41] This platform utilizes mass spectral fragmentation patterns (MS/MS spectra) to build networks of related compounds based on the assumption that structurally similar molecules will fragment in a similar way. A series of molecular networks are built by utilizing the full MS and MS/MS data, and the ions that are structurally related are connected in distinct clusters. GNPS also includes a public database that enables sharing of MS/MS spectra among the scientific community, enabling the mining of spectral data to prioritize samples and screen for chemical moieties of interest.[38, 40-41, 49-53] Using this database, it is possible to putatively identify known natural product secondary metabolites in a given dataset.[40, 52]

CHAPTER II

IDENTIFICATION UTILIZING DATA MINING TECHNIQUES: MASS DEFECT FILTERING AND MOLECULAR NETWORKING

2.1 Abstract

The goal of this research is to compare molecular networking[38, 40, 54] and mass defect filtering[43-44, 55] as ways to screen natural product mixtures for the presence of analogues using the ambuic acid producer *P. microspora* as a case study. With this project, we sought to demonstrate the advantages and disadvantages of both mass defect filtering and molecular networking and provide recommendations to researchers utilizing these technologies for natural products discovery research.

2.2 Materials and Methods

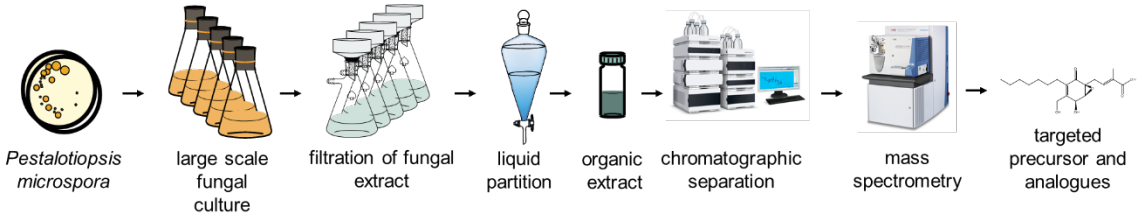


Figure 1. The Workflow for Prioritization of Ambuic Acid Analogues from the *Pestalotiopsis microspora* Organic Extract. *P. microspora* is cultured onto a solid surface agar plate, then inoculated into a liquid medium to be cultured for several weeks. Liquid culture is then filtered and extracted with chloroform and methanol and further separated using liquid-liquid partitioning between water and organic solvents (methanol and acetonitrile). The organic extract is subjected to chromatographic separation and

high- resolution mass spectrometry profiling. Using data mining techniques on the mass spectrometry data can aid the user to target analogues of interest that can subsequently be targeted for isolation.

2.2.1 Solid State Fermentation

P. microspora culture was procured from Dr. Gary Strobel at the University of Montana[32] and grown on solid state oatmeal media. Oatmeal media (Old fashioned Quaker oats) was prepared using 10 g of rolled oats in seven, 250 mL Erlenmeyer flask with approximately 17 mL of deionized-H_2O, followed by autoclaving at 221°C for 30 min using methods outlined previously.[56]

2.2.2 Extraction of Fungal Cultures

To each culture, 60 mL of 1:1 chloroform: methanol ($CHCl_3$:CH_3OH) were added. Solid media cultures were chopped with a spatula to facilitate extraction. The cultures were stirred overnight at room temperature and subsequently filtered under vacuum. The filtrate was stirred for 1 hr with 90 mL of $CHCl_3$ and 100 mL deionized-H_2O. This mixture was then transferred to a separatory funnel, and the bottom organic layer was removed and evaporated *in vacuo*. The dried organic extract was reconstituted with 100 mL 1:1 acetonitrile: methanol and 100 mL hexanes. This biphasic solution was stirred for 1 hr and transferred to a separatory funnel. The bottom layer, representing the defatted organic extract, was evaporated to dryness *in vacuo*.[35]

2.2.3 Data-Dependent UHPLC-PDA-MS/MS Analysis

Fungal extracts were dissolved in methanol to obtain a final concentration of 0.1 mg/mL. Chromatographic separation was performed on an Acquity Ultra-High Performance Liquid Chromatography (UPLC) system (Waters, Milford, MA) equipped with an autosampler kept at 10°C, photodiode array detector (PDA), column manager and binary solvent manager, interfaced to a Q-Exactive Plus mass spectrometer (Thermo Scientific, Bremen, Germany), using an electrospray ionization (ESI) source operated in both negative and positive ion mode. The UHPLC conditions were as follows: Waters BEH C18 100 × 2.1 mm column, 1.7 μm; mobile phase, (A) water with 0.1% formic acid; (B) acetonitrile with 0.1% formic acid; flow rate, 300 μL min^{-1}; injection volume, 3 μL; gradient, linear gradient of 15–85% B over 10 min. The optimized ESI parameters were as follows: source voltage, 3.5 kV (pos); sheath gas flow rate (N_2), 48 units; auxiliary gas flow rate, 11 units; spare gas flow rate, 2.0; capillary temperature, 300°C (pos), S-Lens RF Level, 55. The data-dependent MS/MS events were performed on the three most intense ions detected in full scan MS. The MS/MS isolation window width was 2 Da, and the normalized collision energy (NCE) was set to 40 units. In data-dependent MS/MS experiments, full scans were acquired at a resolution of 35,000 fwhm at a range of *m*/*z* 100 to 1200) and MS/MS scans at 17,500 fwhm, both with a maximum injection time of 50 ms.

2.2.4 Molecular Networking

The .RAW (Thermo) file format of the organic (methanol and acetonitrile) extract was transformed into .mzXML text-based format using the MSConvert software, part of the ProteoWizard package, data is described by the scan number, precursor *m/z* and MS/MS fragmentation information.[57] The converted files were then uploaded to Global Natural Products Social molecular networking (GNPS)[41] for the creation of a molecular network. MS/MS spectra were window filtered by choosing only the top six peaks in the ±50 mDa window throughout the spectrum. The molecular network clusters together nodes using user-defined parameters. The data were clustered with a precursor mass tolerance of 100 mDa and an MS/MS fragment ion tolerance of 0.5 Da to create consensus spectra.[52] A network was then created where edges were filtered to have a cosine score above 0.7 and more than five matched fragment peaks.

2.2.5 Mass Defect Filtering

LC-MS data were collected in negative and positive mode and individually analyzed, aligned, and filtered using MZmine 2.21.2.[58] The .RAW (Thermo) mass spectral data file for the organic (1:1 acetonitrile: methanol) *P. microspora* extract was imported into MZmine for peak picking. The negative polarity has a higher ionization efficiency for the targeted compounds and only masses within the negative peak lists are reported in this study. The chromatogram peak detection was constructed for all *m/z* values lasting longer than 0.1 min, the baseline was set to 3×10^6, *m/z* variation tolerance

at 0.05, and *m/z* intensity and variation tolerance at 20%. The adjusted chromatogram then underwent a deconvolution algorithm to recognize the individual peaks. Peaks were aligned if their masses were within 5 ppm and their retention times were a maximum of 0.15 min from one another. The resulting data matrix, consisting of *m/z*, retention time, and peak area, was imported into Excel (Microsoft). In search for isolatable analogues throughout the chromatogram, this list was filtered for a final time to obtain only *m/z* values that would be in the ±50 mDa mass defect window around the $[M-H]^-$ ion for ambuic acid (*m/z*=349.1651, $C_{19}H_{25}O_6$,) so that the defect range being targeted is 0.1151 to 0.2151. The peak list was sorted by the exact mass, after which these values were truncated into nominal mass and the mass defect. The list was then sorted in reference to descending mass defect, the ions that fell outside of this range were removed.

2.3 Results

2.3.1. Analysis of the *Pestalotiopsis microspora* Fungal Extract

The organic (1:1 methanol and acetonitrile) extract of *P. microspora* was profiled using mass spectrometry and examined for analogue prioritization using mass defect filtering and molecular networking data mining approaches (Figure 1). Ambuic acid was confirmed to be present in the *P. microspora* extract by comparing accurate mass and retention time to that of a standard of ambuic acid (AG-CN2-0129-M001, Adipogen Corp., San Diego, CA). Ambuic acid eluted at 4.67 minutes and is the ion with the highest signal in the *P. microspora* extract (Figure 2). The MS/MS spectra are also

compared as a confirmation, putatively identified ambuic acid ion and that of the standard also showed a high degree of similarity in mass and relative abundance of fragments detected (Figure 2.) MS/MS fragmentation pattern for the associated ambuic acid peak is visualized to the right of the peak in the C and D sections of Figure 2 for the ambuic acid peak in sections A and B, respectively.

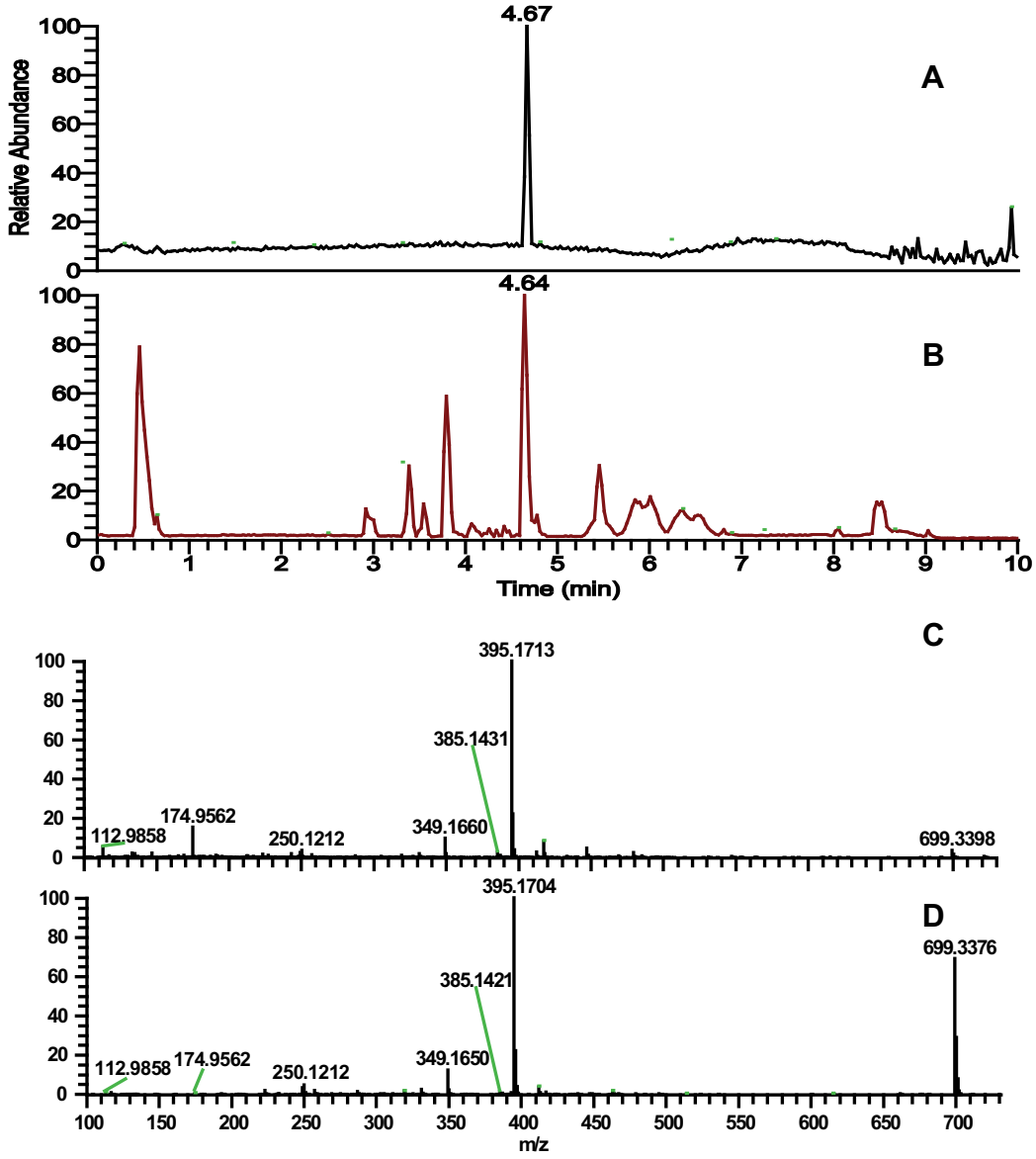
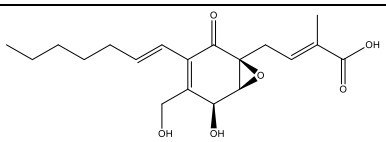
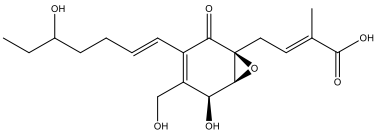


Figure 2. Comparison of Base Peak LC-MS Chromatogram and MS/MS Fragmentation Pattern for Ambuic Acid Detection. (A) Base peak LC-MS chromatogram of an ambuic acid standard analyzed at a concentration of 0.1 mg/mL. (B) Base peak LC-MS chromatogram of an extract, the organic (methanol: acetonitrile) layer of *Pestalotiopsis microspora* culture was analyzed at a concentration of 0.1 mg/mL. (C and D) MS/MS fragmentation pattern for the ambuic acid peak in the A and B sections, respectively.

2.3.2 Mass Defect Filtering Analysis

The mass defect of a given molecule is defined as the deviation between its nominal mass and its exact monoisotopic mass. For example, the $[M-H]^-$ ion of ambuic acid has an exact mass of 349.1651, a nominal mass of 349, and a mass defect of 0.1651 (Table 1). The difference in mass between ambuic acid and an analogue with one additional oxygen atom is quite significant (349.1651 versus 365.1600). However, the mass defects for these two structures (0.1651 versus 0.1600) only differ by 0.0051 Da, representing the mass defect specific to oxygen. Mass defect filtering identifies ions that are within a defined mass defect window around that of the compound of interest based on the assumption that compounds with similar chemical formulas (and thus similar mass defects) will be structurally similar. In this way, it is possible to filter a complex LC-MS chromatogram or large dataset to identify a series of molecules with similar mass defects for the discovery of structurally related components.

Table 1. A Simplified Example of Mass Defect Filtering for the Target Analyte Ambuic Acid and a Hydroxylated Analogue. The difference between the exact and nominal mass is the mass defect. The mass defect will vary depending on the elemental composition of the compound. This example shows an addition of an oxygen that is responsible for the 0.0051 Da difference between the two mass defects.

Name	Structure	Exact mass	Nominal Mass	Mass defect[b]
Ambuic acid ($C_{19}H_{26}O_6$)		349.1651[a]	349	0.1651
Ambuic acid + oxygen ($C_{19}H_{26}O_7$)		365.1600[a]	365	0.1600
Oxygen		15.9949	16	-0.0051

[a] Represent negatively charged ions, $[M-H]^-$ in the mass spectrometer
[b] Exact mass- nominal mass = mass defect

The goal of this study was to assess the ability of mass defect filtering and molecular networking to identify analogues of ambuic acid for natural product discovery, only ions detected above the baseline of 3 x 10^6 were considered for analysis. Of the 183 ions detected above baseline in negative ionization mode (Figure 2), 84 ions had mass defects within 50 mDa from that of ambuic acid (mass defect range of 0.1151 to 0.2151).

2.3.3 Molecular Networking Analysis

A visual explanation of how data is processed for molecular networking is represented in Figure 3. The full MS spectrum (A) is utilized to create the final molecular

networks for analysis. Before the networks are formed, connections between nodes are created computationally within the GNPS platform by converting the MS/MS spectra into vectors and comparing their similarities (B). The vector units undergo a pair-wise dot product calculation to calculate the cosine similarity score where higher values (maximum of 1) represent similar fragmentation patterns(C).[52] Cosine similarity scores are used to plot the final molecular networks (D), in which each circular node in the final molecular network represents an *m/z* value of an ion detected in the full MS chromatogram. A network is a group of nodes, which represent ions tat fragment similarly.

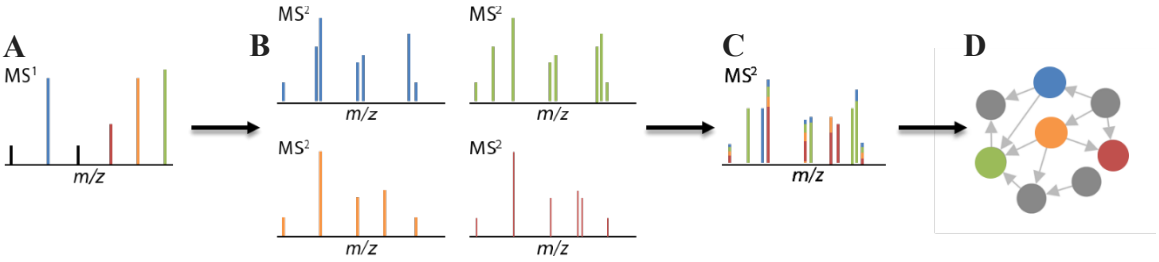


Figure 3. A Simplified Visual Representation of Global Natural Product Social Molecular Networking (GNPS).[41] (A) Full MS spectra are acquired. (B) MS/MS data are utilized to identify clusters of m/z values that correspond to structurally related compounds. (C) Connectivity or relativeness of the nodes is calculated computationally and (D) the results are visually represented by a molecular network. The circular nodes connected in a final molecular network are representative of ions detected for which MS/MS spectra were similar based on computed similarity scores.

Molecular networks were produced using MS/MS data from the *P. microspora* organic extract (Figure 2). One molecular network contained accurate masses corresponding to ambuic acid and several associated adducts (ambuic acid dimer, formic

acid adduct, and chlorine adduct) along with putative ambuic acid analogues (Figure 4). The cluster of interest contained 17 ions that were above the cosine similarity scoring of 0.7 and shared at least 5 MS/MS fragments with ambuic acid. In this molecular network, green nodes represent *m/z* values specific to ambuic acid, red nodes represent *m/z* values that were found using both mass defect filtering and molecular networking, and the blue nodes were only found using molecular networking. Notably, molecular networking identified a much smaller set of ions related to ambuic acid than was obtained with mass defect filtering. While mass defect filtering identified a subset of 84 putative analogues of ambuic acid, the molecular network is comprised of only 17 ions.

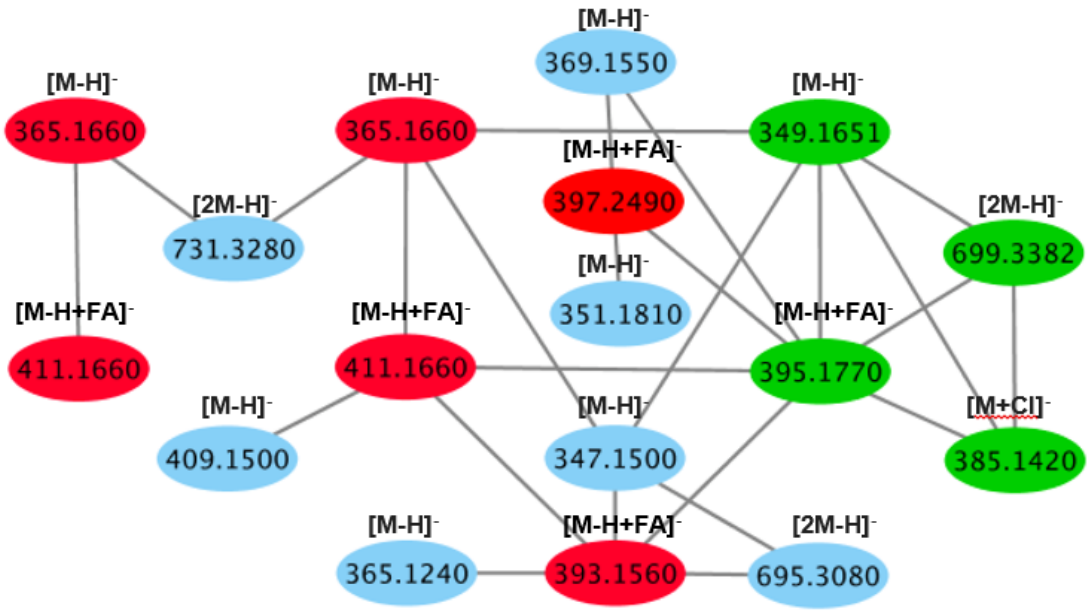


Figure 4. Molecular Network Containing Ambuic Acid and Putative Structural Analogues. The network is composed of the ions that have similar MS/MS fragmentation pattern to ambuic acid and adducts of ions are labeled with appropriate ion description: $[M-H]^-$ (deprotonated molecular ion), $[M-H+FA]^-$ (formic acid adduct), $[2M-H]^-$ (proton bound dimer of two deprotonated molecular ions), and $[M+Cl]^-$ (chloride adduct). The green nodes represent ambuic acid and mass spectral adducts. The red nodes represent

ions that were identified as potential ambuic acid analogues by both mass defect filtering and molecular networking, and blue nodes were identified as potential analogues only by molecular networking.

2.3.4. Comparison of Molecular Networking and Mass Defect Filtering Analysis of *P. microspora* Data

All priority peaks were compiled into one master list, shown in Table 2. Each ion was manually examined in the. RAW (Thermo) file on XCalibur and noted if there was fragmentation of this ion not. Each ion is annotated with a footnote on discovery method. (MN=molecular networking; MDF=mass defect filtering; a=MDF only, no fragmentation; b=MDF only, fragmentation patterns do not match; c= in both, MDF and MN; d=MN only, below baseline; e=MN only, outside MDF window) There were five main ion categories that were created and more discussion is available in Figure 5. The ion number were also utilized in Table 3, when the large peak list was further filtered to identify peaks that were found to be related to ambuic acid.

Table 2. Mass Defect Filtered and Molecular Networking Peak List. The. RAW (Thermo) data is converted to mzML format and imported onto GNPS for molecular networking. The resulting cluster analysis is searched for the node represented for ambuic acid m/z ions. All ions for with the m/z falls within ±50 Da window of ambuic acid ions ($2M-H]^-$, $[M+FA-H]^-$, $[M-H]^-$) are included. $[M-H]^-$, $[M+FA-H]^-$, $[2M-H]^-$, and $[M+Cl-H]^-$ represent negatively charged ions produced by mass spectrometer in negative ion mode and mean the precursor exact mass minus a proton, precursor exact mass with a formic acid (FA) adduct minus a proton, double the precursor exact mass minus a proton, and precursor exact mass with chlorine (Cl) adduct minus a proton, respectively.

Ion number	*m/z* value	Retention time	Fragmented?	Discovery method
1	223.1335	4.64	N	a
2	250.1205	4.64	N	a
3	257.1543	4.64	N	a
4	265.1475	8.52	N	a
5	265.1476	5.46	N	a
6	265.1476	8.05	N	a
7	266.1155	3.38	N	a
8	266.1508	5.46	N	a
9	267.1419	5.46	N	a
10	287.1647	4.64	N	a
11	293.1790	6.42	Y	b
12	297.1527	5.4	Y	b
13	298.1559	5.48	Y	b
14	307.1947	6.98	Y	b
15	311.1683	8.47	Y	b
16	311.1685	8.05	Y	b
17	311.1685	6.01	Y	b
18	312.1716	6.01	N	a
19	319.1546	4.64	N	a
20	325.1839	8.52	Y	b
21	325.1840	6.36	Y	b
22	325.1841	8.05	Y	b
23	326.1872	6.36	N	a
24	327.1404	0.46	Y	b
25	328.1241	0.46	N	a
26	331.1547	4.64	N	a
27	339.1996	6.83	Y	b
28	347.1500	4.33	Y	d
29	349.1651	4.64	Y	c
30	349.1655	4.12	Y	b

31	350.1685	4.64	N	a
32	351.1808	3.79	Y	b
33	351.1810	3.55	Y	b
34	351.1810	4.77	Y	d
35	352.1842	3.79	N	a
36	355.1578	8.47	Y	b
37	365.1240	3.51	Y	d
38	365.1600	2.97	Y	c
39	365.1602	3.36	Y	c
40	366.1633	3.38	N	a
41	366.1634	2.92	N	a
42	367.1758	3.44	Y	b
43	369.1550	3.91	Y	d
44	377.1965	6.33	N	a
45	377.1966	6.12	N	a
46	379.1579	8.05	N	a
47	381.1735	8.52	Y	b
48	382.1758	8.52	N	a
49	383.1710	3.11	Y	b
50	385.1420	4.61	Y	d
51	385.1422	4.64	Y	c
52	387.1577	3.79	N	a
53	388.1176	0.43	N	a
54	391.1761	4.26	Y	b
55	393.1555	4.48	Y	c
56	393.1918	4.07	N	a
57	394.1949	4.07	N	a
58	395.1707	4.23	Y	b
59	395.1708	4.64	Y	c
60	396.1738	4.64	N	a
61	397.1500	3.41	Y	b
62	397.1758	4.64	N	a
63	397.1863	3.55	N	a
64	397.1864	4.78	N	a
65	397.1864	3.79	N	a
66	397.2490	4.24	Y	e
67	398.1895	3.79	N	a
68	398.1897	3.55	N	a
69	398.1898	4.78	N	a
70	399.1915	3.79	N	a
71	409.1500	3.02	Y	d

72	411.1657	2.97	Y	c
73	411.1657	3.36	Y	c
74	411.1659	3.22	Y	b
75	411.1659	3.9	Y	b
76	411.1659	4.42	Y	b
77	411.2019	3.77	N	a
78	412.1607	4.64	Y	b
79	412.1690	3.38	N	a
80	412.1690	2.92	N	a
81	413.1815	3.74	N	a
82	413.1815	3.33	N	a
83	414.1764	3.79	Y	b
84	417.1524	4.61	N	a
85	419.1682	3.82	N	a
86	427.1604	3.44	Y	b
87	427.1604	3.52	Y	b
88	429.1764	3.11	Y	b
89	433.1472	3.38	N	a
90	463.1579	4.64	N	a
91	465.1737	3.79	N	a
92	695.3090	4.38	Y	e
93	699.3380	4.61	Y	e
94	731.3280	3.36	Y	e
MN=molecular networking; MDF=mass defect filtering; a=MDF only, no fragmentation; b=MDF only, fragmentation patterns do not match; c= in both, MDF and MN; d=MN only, below baseline; e=MN only, outside MDF window				

Having multiple data mining methods to examine data and prioritize sample components is helpful to ensure a complete interpretation of the dataset. A combined list of possible ambuic acid analogues (94 prioritized ions) identified by mass defect filtering and molecular networking are provided in Table S1. Some ions were only identified by one data analysis approach, while others were identified by both methods. Figure 5 is a visual depiction of the distribution of the 94 compiled prioritized peaks. A total of 183 ions were detected above the baseline of 3 × 106 in the negative ion mode, 84 of which

had a mass defect within 50 mDa of that of ambuic acid. This is a lengthy list, representing nearly half of all ions detected, and isolation of all these ions to confirm their identity would not be possible given constraints of time and sample quantity. Thus, we sought to further narrow the list of putative analogues by employing molecular networking.

Out of the 84 ions prioritized by mass defect filtering, 46 ions (orange) were not fragmented using the data-dependent LC-MS/MS method. This could occur due to their low abundance or because of co-elution with more abundant ions. There were 30 ions from among those prioritized with mass defect filtering (green) that had fragmentation patterns that did not match ambuic acid. These ions were likely falsely identified as ambuic acid analogues by mass defect filtering, illustrating that compounds with similar mass defect do not necessarily share structural relatedness. There were 8 ions (blue) that fell within the mass defect window of interest and shared similar fragmentation patterns with ambuic acid. An additional 10 ions were identified only by molecular networking, 6 of which (red) were not identified by mass defect filtering because they were detected below the baseline used for peak picking (3×106), and 4 ions (purple) fell above the mass defect window of ±50 mDa. Notably, these 3 of the 4 ions outside the mass defect range represent dimers $[2M-H]^-$ of analogues, and one represented an unknown. When looking at the data via the mass defect filtering, these ions would be considered a false negative, because they were not included in the priority list. However even though they were outside the range of the mass defect window, when looking back at the .RAW data

the discovery of the dimers would be apparent. A $[2M-H]^-$ dimer is commonly seen in mass spectrometry[59] and the peak list would contain the precursor ion. Having dimers in the peak list would be a secondary ion for one component. The fourth ion in this category was an unknown and was a false positive that would not have been discovered without the secondary data mining technique. The fragmentation fingerprint was examined to determine structural similarity along with other examples of ions in each of the categories.

Having this comparison illustrates that while there is some overlap between mass defect filtering and molecular networking (8 ions), the ions identified as putative analogues are largely different using the two different approaches. Mass defect filtering provided a much larger list, containing many putative false positives (30 out of 84). Molecular networking has prioritized a much smaller subset of ions but could risk false negative results by only looking at ions with MS/MS spectra. Over half of the ions identified by mass defect filtering were not fragmented using the data-dependent LC-MS method and cannot be confirmed or rejected as putative analogues.

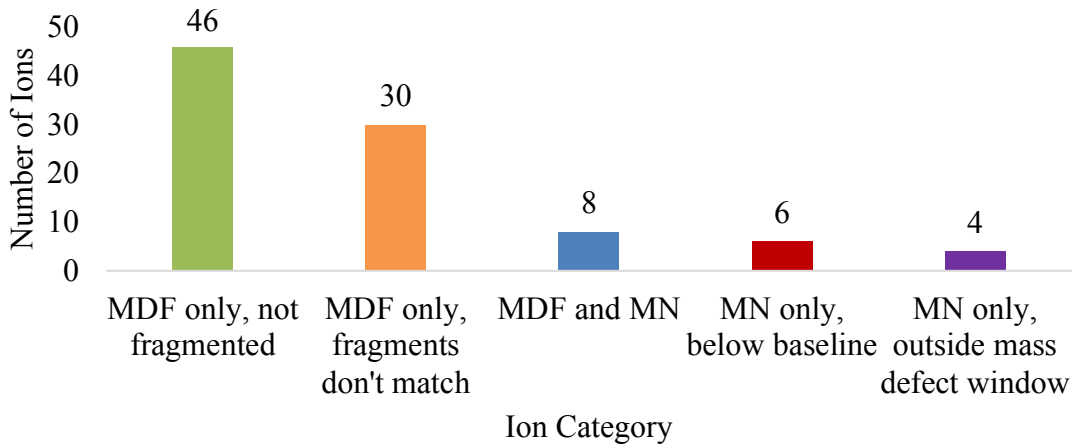


Figure 5. Comparison and Categorization of Ions Detected with Mass Defect Filtering (MDF) and Molecular Networking (MN). After mass defect filtering, 84 ions were prioritized, 46 (orange) were not fragmented and the structural relatedness could not be determined, 30 (green) had fragmentation but did not match the target analyte, and 8 ions (blue) overlapped with those identified using molecular networking. There were 10 additional ions identified by molecular networking alone, in which 6 (red) were below the threshold baseline set in the mass defect filtering and 4 (purple) were outside the mass defect window of 50 mDa from ambuic acid. 3 out of the 4 (purple) were due to dimers, and one was and unknown that has a fragmentation pattern that is similar to ambuic acid.

Fragmentation patterns of selected ions identified as putative ambuic acid analogues by one or both are illustrated in Figure 6. Fragmentation spectra of the analyte of interest, ambuic acid, are depicted in Figure 6A (the most abundant precursor ion of ambuic acid is the formic acid adduct $[M+FA-H]^-$, *m/z* 395.1708) and fragments that were found in the spectra for putative analogues have been highlighted and annotated with corresponding *m/z* values. In Figure 6B, the fragmentation spectrum of a putative analogue identified by both mass defect filtering and molecular networking are shown. This compound *(m/z* 365.1602) shares many fragments with ambuic acid and is likely structurally related. In Figure 6C, fragmentation spectrum of an analogue only identified

with molecular networking and was a false-negative for mass defect filtering is shown (*m/z* 731.3280). Notably, the fragmentation spectra still show a high degree of overlap with the spectra of ambuic acid (Figure 6A), even though the mass defect (0.3380) falls outside of the 50 mDa range used for mass defect filtering. Thus, the ion at *m/z* 731.3280 is likely a relative of ambuic acid that would have been overlooked using just mass defect filtering. The example in Figure 6D is an ion (*m/z* 397.2228) that was only found by molecular networking and was outside of the mass defect range. This was the only ion in this category that was not a dimer and is a true false negative. This ion would have not been seen through data processing if mass defect filtering was the only mining technique. Finally, Figure 6E illustrates the fragmentation spectrum of a compound identified only by mass defect filtering (*m/z* 391.1761) and is a representative example of a false-positive of mass defect filtering. Although this compound falls within the mass defect range of ambuic acid, the fragmentation pattern does not overlap with that of ambuic acid, suggesting that it is not structurally related to this compound and that it was falsely identified as an analogue using mass defect filtering.

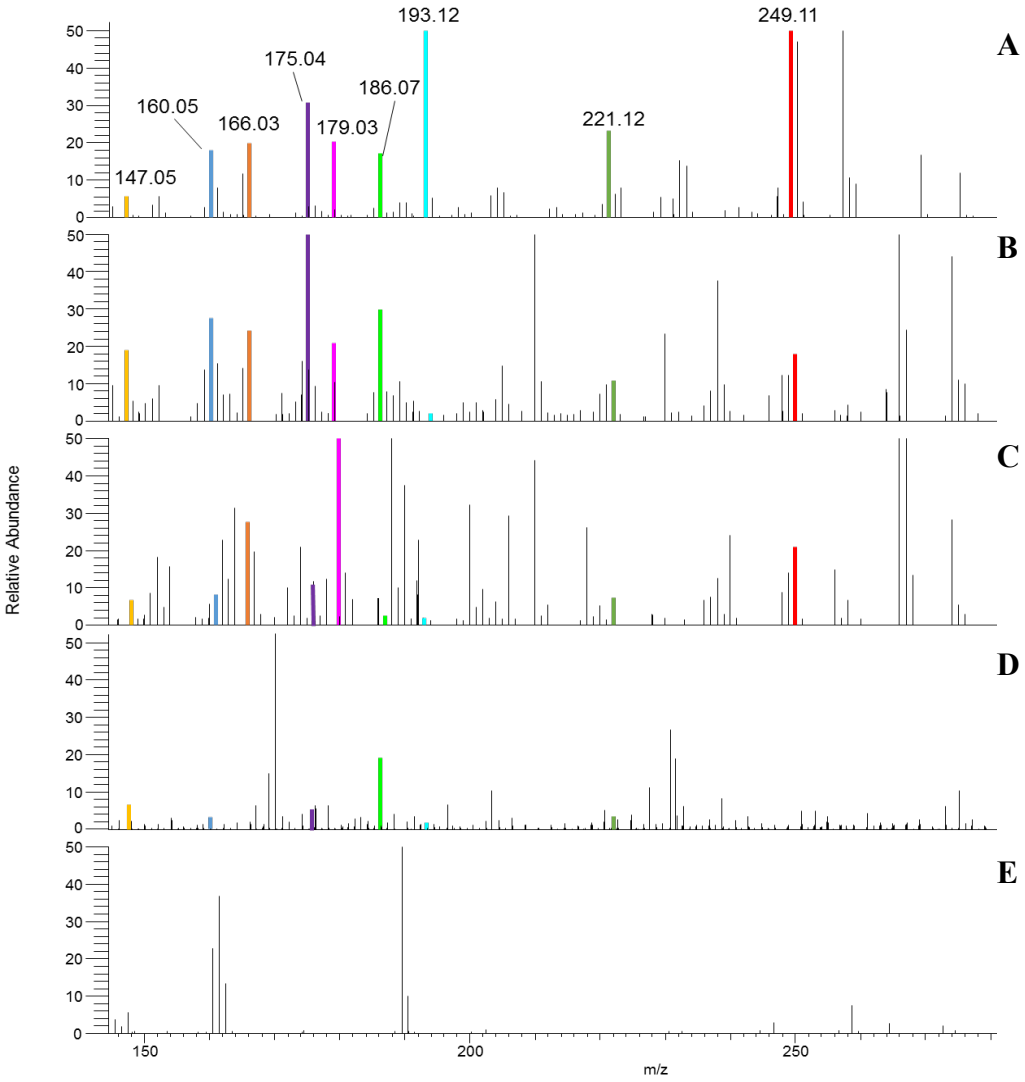


Figure 6. Comparison of the Fragmentation (MS/MS) Spectra of the Putative Ambuic Acid Analogues. For ease of comparison, the main fragments that match in multiple spectra are color coded and labeled with fragment *m/z* value in each spectrum (A-E) if present. (A) Ambuic acid with formic acid adduct in negative mode, $[M-H+FA]^-$ (*m/z* 395.1758), fragmentation pattern. (B) One of the putative analogues discovered by mass defect filtering and molecular networking (*m/z* 365.1600) with highlighted fragments. (C) One of the dimers found only in molecular networking because it was outside of the mass defect range (*m/z* 731.3280) with highlighted fragments. (D) One of the putative analogues found only in molecular networking (*m/z* 397.2228), which has over five fragments required to be clustered with ambuic acid in the molecular network. (E) One of the putative analogues found with mass defect filtering (*m/z* 391.1761), which does not share any fragments with spectrum obtained for ambuic acid.

The way in which the mass defect window is defined directly affects the number of ions that will be identified as putative analogues. A narrower window will result in fewer ions being identified as potential analogues and will minimize the chance of false positives. However, a narrower window will also result in an increase in the number of false negatives. To demonstrate this concept, a comparison of the distribution of ions from the *P. microspora* extract falling in various mass defect ranges is shown in Figure 7. While a total of 84 ions are identified within the mass defect window of 50 mDa (blue bars), only 61 are identified in the 20 mDa mass range. The number of ions tentatively identified as ambuic acid analogues with molecular networking in each given mass window are also shown. The data for analogues identified with molecular networking (orange bars) demonstrates that each time the mass window is narrowed, a greater number of potential analogues is missed due to relevant ions falling outside the chosen window. 3 of the 4 ions in orange in the above 50 mDa category are dimers, the last ion is a putative analogue of ambuic acid and would have been missed. Having the widest window possible would give you the full peak list, the concept of data mining is to filter or funnel the data to prioritize the ions that are indeed related to the desired precursor ion. The user has the ability to define the window in a way that would best benefit the project and would have to be optimized for the goals attempting to achieve.

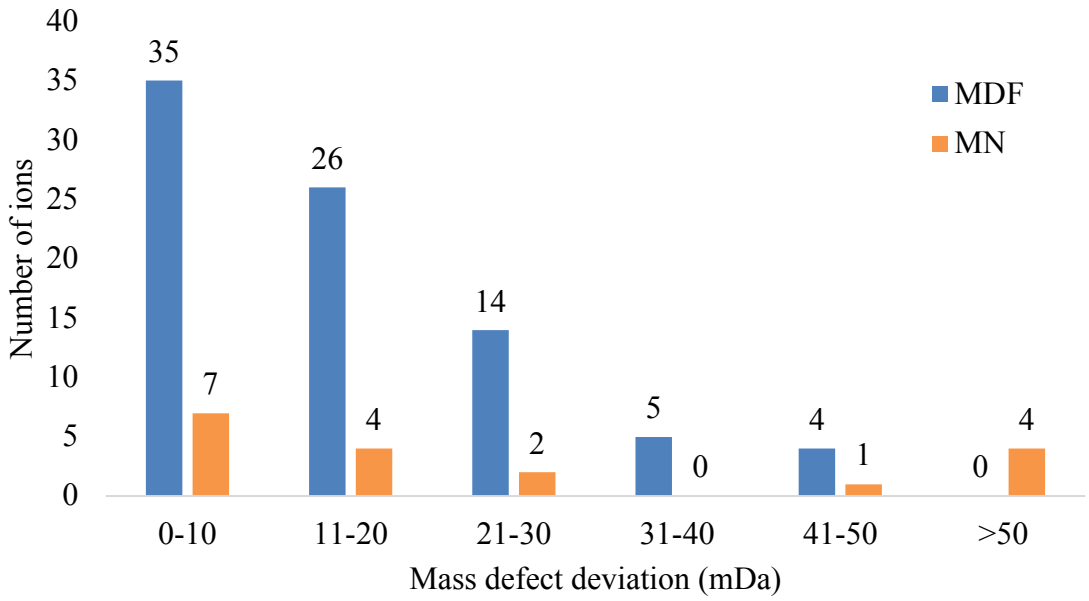


Figure 7. A Comparison of the Distribution of the 94 Prioritized Ions' Mass Defect Deviation(mDa) from Ambuic Acid. The number of ions identified in the given mass defect range using mass defect filtering (MDF, blue) and molecular networking (MN, orange) are shown. The ions in orange have been confirmed by MS/MS data as ions that are identified to be related to ambuic acid. Some ions are repetitious for ions that were more abundant, 3 of the 4 ions in the above 50mDa range is due to dimerization within the mass spectrometer.

Figure 5 had shown the overlap of discovery with both techniques as well as the false-negatives mass defect filtering had left out. 17 ions have been described from the molecular network and have been manually examined to ensure the relatedness, some ions are repetitious for the same component. This same network (Figure 4) has been recolored nodes to visualize the eight unknowns and ambuic acid in Figure 8. The purple nodes are the *m/z* values of Unknown A, green for ambuic acid, red for Unknown B, pink for Unknown C, aqua for Unknown D, yellow for Unknown E, dark blue for Unknown F,

and light green for Unknown G. Unknown D and E are isomers of each other. The nodes have also been labelled with the type of ion and Table 3 summarizes the network.

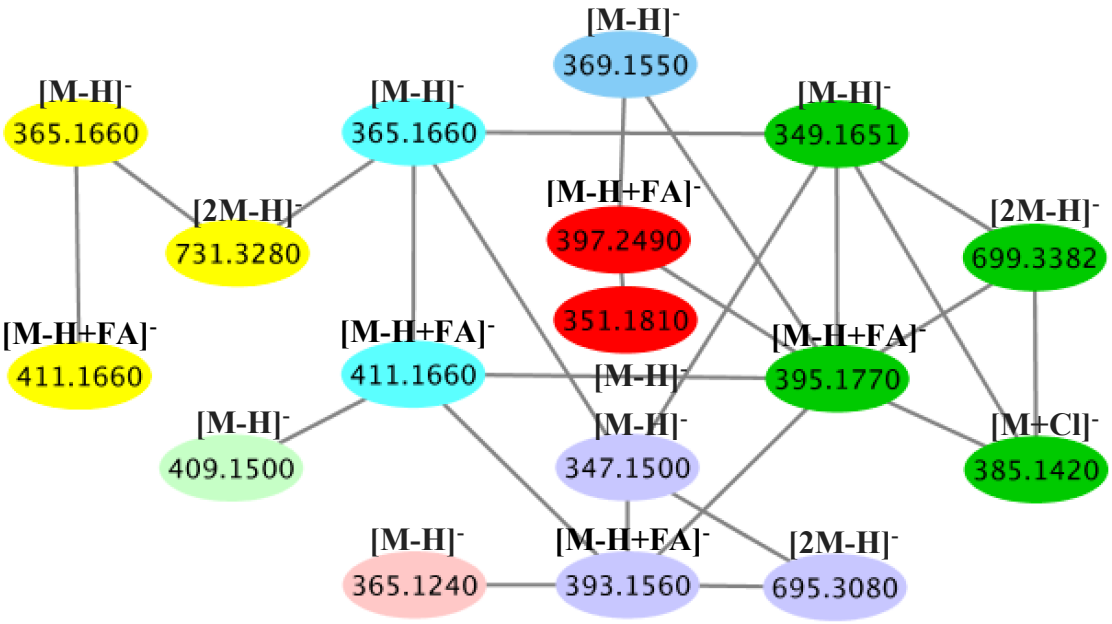


Figure 8. Molecular Network Color-coded into Different Compound Assignments. The network is composed of the ions that have similar MS/MS fragmentation patterns. The adducts of ions are labeled with appropriate description: $[M-H]^-$, $[M-H+FA]^-$, $[2M-H]^-$, and $[M+Cl]^-$, where FA represents formic acid and Cl represents chlorine. The network has identified 8 different compounds within the extract, 1 accounts for ambuic acid and the other 7 other unknowns. The purple nodes are the *m/z* values of Unknown A, green for ambuic acid, red for Unknown B, pink for Unknown C, aqua for Unknown D, yellow for Unknown E, dark blue for Unknown F, and light green for Unknown G. Unknown D and E are isomers of each other.

The has been also transformed into Table 3 with the putative molecular formula. This list is much more manageable for isolation efforts verses the 183 isolatable ions first prioritized. Having the MS/MS data of the ions allows us to make predictions of the putative molecular formula. Confirmation of this would have to be confirmed via

structure elucidation and was out of the scope of this data mining comparison. Table 3 also contains the ion number assigned in original combined peak list for both molecular networking and mass defect filtering in Table 2.

Table 3. Predicted Analogue Identification. The 17 ions visualized by the molecular network are annotated to assessed within the original data file to confirm there were 7 unknowns compounds that have been prioritized with both molecular networking and mass defect filtering to be related to the target analyte, ambuic acid. Unknown D and E are isomers of each other.

Compound	Proposed chemical formula	*m/z* value	Ion Type	Ion number (Table S1)	Molecular Network node color
Unknown A	$C_{21}H_{29}O_7$	347.1500	$[M-H]^-$	28	
		393.1555	$[M-H+FA]^-$	55	
		695.3080	$[2M-H]^-$	92	
Ambuic Acid	$C_{19}H_{26}O_6$	349.1651	$[M-H]^-$	29	
		385.1422	$[M+Cl]^-$	51	
		395.1708	$[M-H+FA]^-$	59	
		699.3383	$[2M-H]^-$	93	
Unknown B	$C_{19}H_{27}O_6$	351.1810	$[M-H]^-$	34	
Unknown C	$C_{18}H_{21}O_8$	365.1240	$[M-H]^-$	37	
Unknown D	$C_{19}H_{26}O_7$	365.1600	$[M-H]^-$	38	
		411.1657	$[M-H+FA]^-$	72	
Unknown E	$C_{19}H_{26}O_7$	365.1602	$[M-H]^-$	39	
		411.1657	$[M-H+FA]^-$	73	
		731.3280	$[2M-H]^-$	94	
Unknown F	$C_{18}H_{25}O_8$	369.1550	$[M-H]^-$	43	
Unknown G	$C_{18}H_{21}O_8$	409.1500	$[M-H]^-$	71	
Unknown H	$C_{21}H_{33}O_7$	397.2240	$[M-H]^-$	66	

All of the 94 prioritized peaks (Table 2) were individually examined for fragmentation data and the final prioritized eight unknowns and ambuic acid first discussed in the molecular network (Figure 8) and the ions have been further described to

connect all of the results from both data mining processes into Table 3. In Figures 9-17 the extracted ion chromatogram of each are each of the related unknowns (Unknown A-H) and ambuic in section A. within section B of Figures 9-17 the fragmentation fingerprint is displayed. On each of the spectra there are the highlighted ions that were detected by molecular network. Majority of the extracted ion chromatograms have multiple peaks throughout the chromatogram and were individually examined to determine whether they are in fact not related to ambuic acid by MS/MS comparison. The retention time is labelled within section a to signify the ion that has been prioritized. It is notable to again reference the discovery of structurally related isomer seen in Figure 13 and 14 for Unknown D and E respectively. To complete the assessment of the analogues, they would have to be individually isolated and subjugated to structure elucidation.

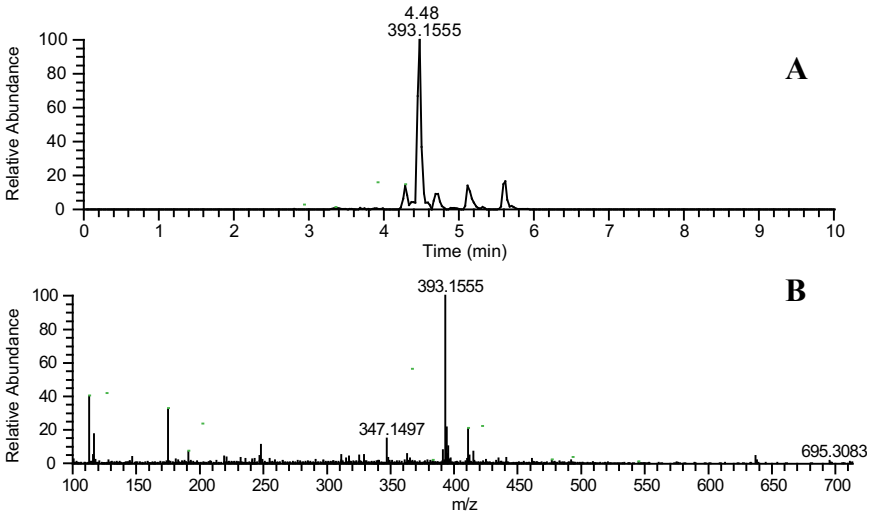


Figure 9. Unknown A Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for Unknown A eluting at 4.48 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 347.1500, 393.1555, 695.3080 representing $[M-H]^-$, $[M-H+FA]^-$, $[2M-H]^-$ respectively.

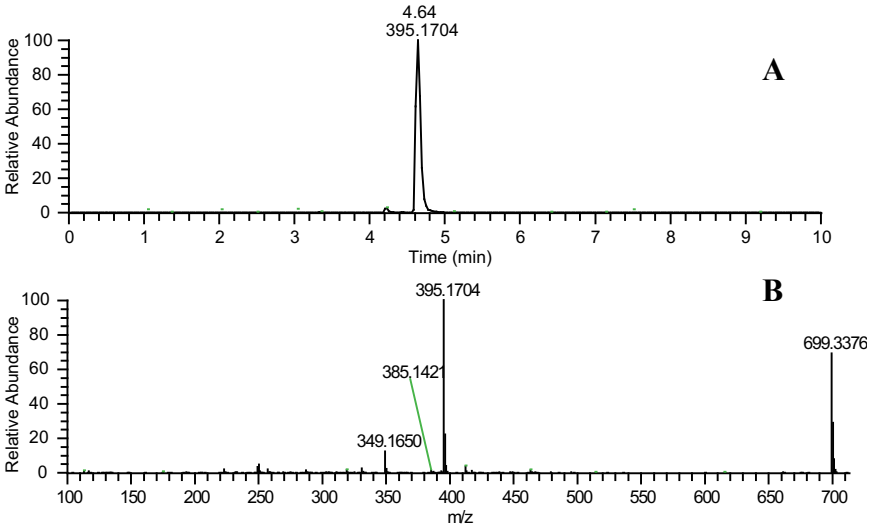


Figure 10. Ambuic Acid Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for ambuic acid eluting at 4.64 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 349.1650, 385.1421, 395.1704, 699.3376 representing $[M-H]^-$, $[M+Cl]^-$, $[M-H+FA]^-$, $[2M-H]^-$ respectively.

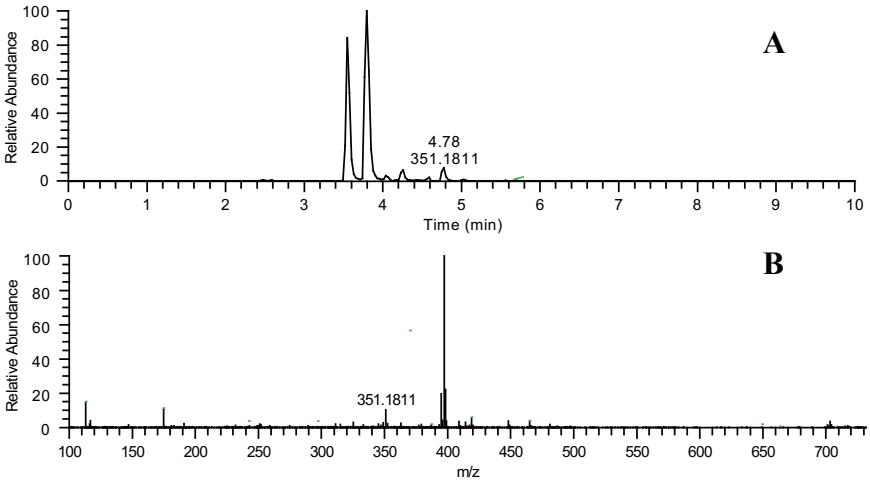


Figure 11. Unknown B Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for ambuic acid eluting at 4.78 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 351.1811 representing $[M-H]^-$.

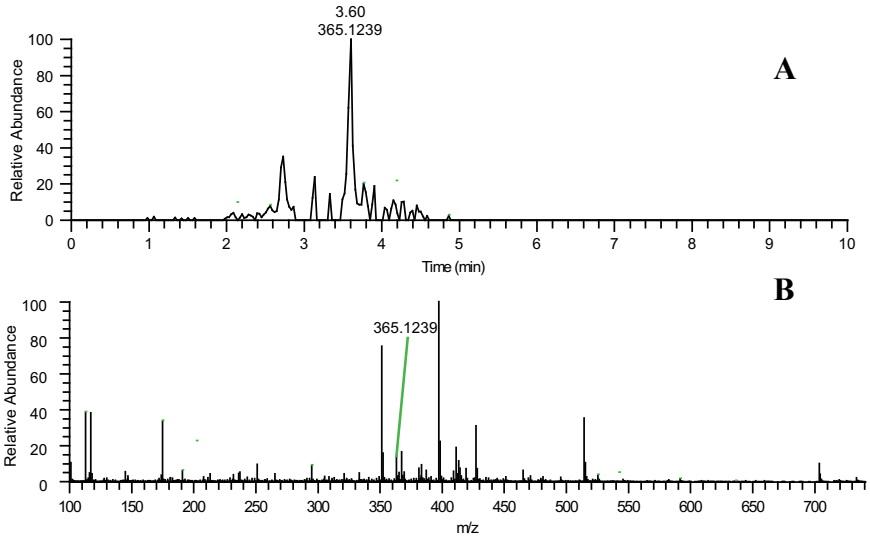


Figure 12. Unknown C Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for ambuic acid eluting at 3.60 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 365.1239 representing $[M-H]^-$.

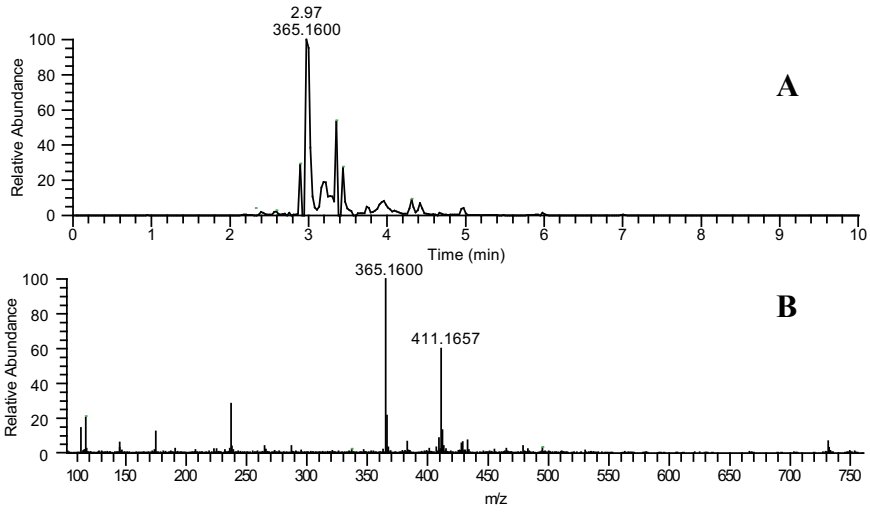


Figure 13. Unknown D Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for ambuic acid eluting at 2.97 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 365.1600 and 411.1657 representing $[M-H]^-$ and $[M-H+FA]^-$ respectively.

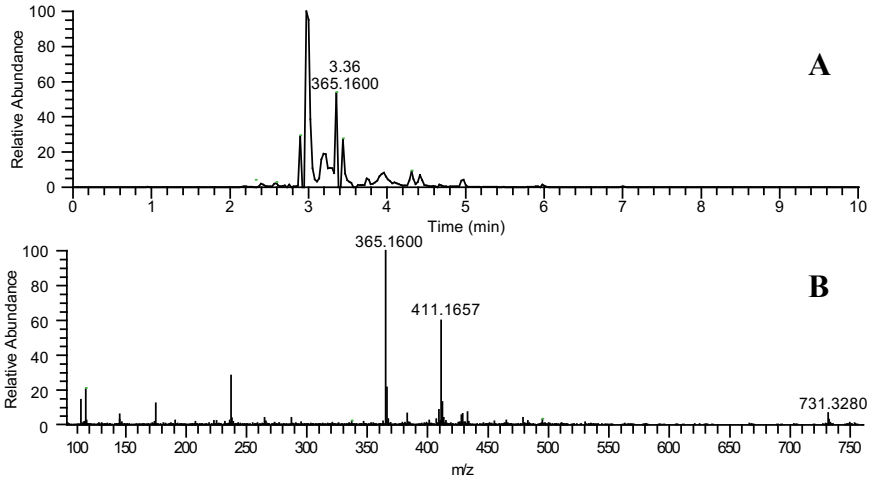


Figure 14. Unknown E Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for Unknown A eluting at 3.36 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 365.1600, 411.1657, 731.3280 representing $[M-H]^-$, $[M-H+FA]^-$, $[2M-H]^-$ respectively.

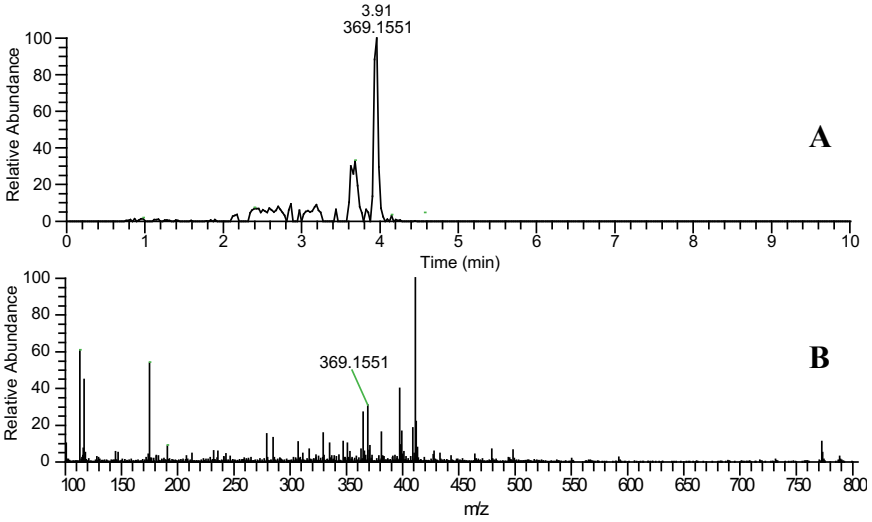


Figure 15. Unknown F Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for ambuic acid eluting at 3.91 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 369.1551 representing $[M-H]^-$.

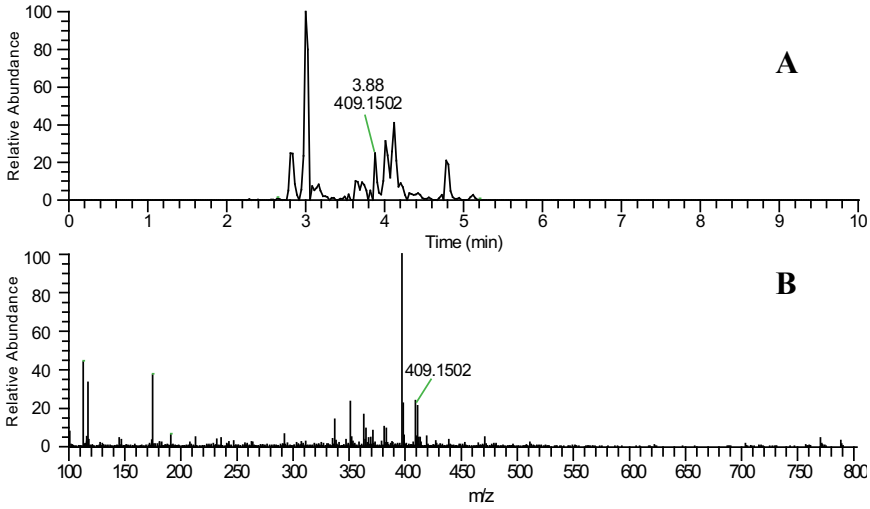


Figure 16. Unknown G Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for ambuic acid eluting at 3.88 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 409.1502 representing $[M-H]^-$.

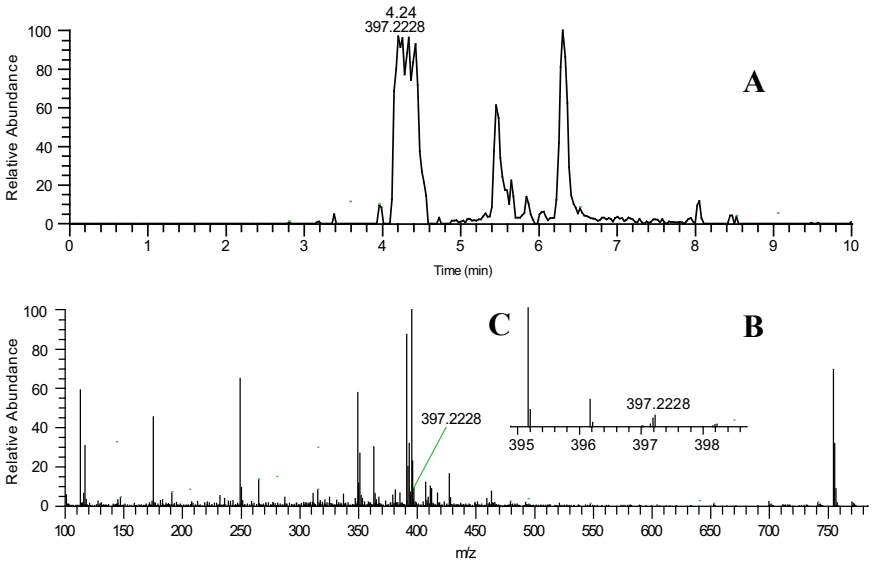


Figure 17. Unknown H Mass Spectral Data. (A) Extracted Ion Chromatogram (XIC) for ambuic acid eluting at 4.24 min. (B) MS/MS fragmentation fingerprint. Highlighted are the ions that were found in the molecular network *m/z* values of 397.2228 representing $[M-H]^-$. (C) A zoom in to 397.2228, which was one of the minor eluting peaks at that time point.

2.4 Discussion

As this study demonstrates, mass defect filtering and molecular networking can be utilized to identify compounds that are likely to be analogues of a molecule of interest. With this project, we aimed to compare techniques (mass defect filtering and molecular networking) to answer the same question: does the complex extract prepared from *Pestalotiopsis microspora* contain analogues of ambuic acid? Interestingly, some putative analogues were only identified with one methodology, while others were identified with both approaches. The comparison of these approaches has allowed us to identify the strengths and weaknesses of each approach.

Mass defect filtering is a simple technique to narrow down full peak lists for peaks of interest using only full MS data. Because mass is an intrinsic property of molecules, it does not change from one analysis to the next. As such, the mass defect, extracted from the finite exact mass of a desired precursor ion, can be used to interpret the elemental components of a molecule, as well as to identify other compounds within a complex mixture with similar elemental composition. The simplicity of this method offers an advantage for laboratories with limited access to MS/MS capable instrumentation, where mass defect filtering can serve as an excellent starting point to assess whether putative analogues of a desired compound of interest exist within a given sample. The subsequent peak list can be used to build a list of ions to prioritize for follow up studies. By including the generation of MS/MS spectra of prioritized ions, enabling

direct comparison of fragmentation patterns to determine which of the ions are, in fact, related.

Despite its accessibility and simplicity, the mass defect filtering approach is more prone to false positives than is molecular networking. Without MS/MS fragmentation patterns, it is challenging to confirm whether each result from the mass defect filtering is structurally related. When applied to the simplified *P. microspora* case, mass defect filtering identified a subset of 84 masses for prioritization for the one extract. Clearly, the isolation of 84 ions to confirm structure is not feasible. However, it would be possible to manually examine the MS/MS spectra to determine which among them are likely to be ambuic acid analogues as a way to prioritize compounds for isolation. The molecular networking dataset yielded only 17 ions, which is a more tractable list for follow up isolation studies. Out of the 84 ions that were prioritized in this simplified example (representing nearly half of all ions detected above baseline), nearly 35% did not possess similar fragmentation spectra to ambuic acid (Figure 6). An additional 46 ions were not fragmented using the LC-MS method, and it was not possible to confirm whether these are related to ambuic acid. These ions may have had a low signal intensity compared to other ions detected at each scan or may have co-eluted and undergone ion suppression with multiple high abundant analytes, and as such may not have been fragmented with the data-dependent method utilized for this study.

Molecular networking has advantages with untargeted analysis as well as the targeted identification of known compounds. This approach has the ability to add

annotations from multiple small molecule databases by matching MS/MS spectra within the user's data to other spectra uploaded on the website.

The molecular network clusters together nodes using user-defined parameters. Molecular networking, unlike mass defect filtering, involves numerous steps and data transformations and could be prone to misinterpretation if parameters are not optimized. However, when optimized parameters are used, the approach can be highly effective for identifying analogues of the molecule of interest. There are some notable disadvantages of molecular networking. The molecular networking approach is far more complicated as a method of data analysis, and it *only* identifies analogues for which MS/MS data are generated. If an extract component is not detected at sufficient abundance to generate an MS/MS spectrum, it will not be detected with molecular networking. Using the mass defect information of the 46 ions that were not fragmented, for complete coverage of the metabolome adding this to an inclusion list when acquiring mass spectrometry data can confirm whether there is relatedness.

Finally, it is worth noting that several possible analogues were only detected with either mass defect filtering or molecular networking (Figure 5), thus, to achieve the most comprehensive data mining interpretation, it would be useful to employ both techniques. A molecular network has been color-coded into compound assignment (Figure 8), the consolidation of data and predicted analogue identification in Table 3, and the confirming MS/MS spectra for ambuic acid and the eight unknowns and their associated adducts in Figure 9-18.

2.5 Conclusion

The range of metabolites produced by fungal organisms such as *P. microspora* is vast, and analysis of such complex mixtures by LC-MS can yield enormous datasets. The results of this study illustrate that post-acquisition filtering of these datasets, including mass defect filtering and molecular networking, can facilitate target identification by removing signals that are not related to the metabolite of interest. Mass defect filtering and molecular networking represent complementary approaches to answer the same question and can be used in tandem to group structurally related compounds in complex natural product mixtures. Ultimately, these approaches both have unique benefits and utilization of both approaches represents a promising strategy for natural product lead prioritization.

REFERENCES

1. CDC Antibiotic Resistance Threats in the United States, 2013; CDC: Atlanta, GA, **2013**; pp 1-114.
2. CDC. Antibiotic Use in the United States, 2017: Progress and Opportunities. Atlanta, GA: US Department of Health and Human Services, CDC. **2017**.
3. Rice, L. B., Federal funding for the study of antimicrobial resistance in nosocomial pathogens: no ESKAPE. *J Infect Dis.* **2008,** *197* (8), 1079-1081.
4. Infectious Diseases Society of, A., The 10 x '20 Initiative: pursuing a global commitment to develop 10 new antibacterial drugs by 2020. *Clin Infect Dis.* **2010,** *50* (8), 1081-1083.
5. Boucher, H. W.; Talbot, G. H.; Bradley, J. S.; Edwards, J. E.; Gilbert, D.; Rice, L. B.; Scheld, M.; Spellberg, B.; Bartlett, J., Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis.* **2009,** *48* (1), 1-12.
6. World Health Organization. The evolving threat of antimicrobial resistance : options for action. Geneva : World Health Organization. **2012.**
7. Fischbach, M. A.; Walsh, C. T., Antibiotics for emerging pathogens. *Science.* **2009,** *325* (5944), 1089-93.
8. Sukumaran, V.; Senanayake, S., Bacterial skin and soft tissue infections. *Aust Prescr.* **2016,** *39* (5), 159-163.
9. David, M. Z.; Daum, R. S., Community-associated methicillin-resistant Staphylococcus aureus: epidemiology and clinical consequences of an emerging epidemic. *Clin Microbiol Rev.* **2010,** *23* (3), 616-687.
10. Farley, J. E., Epidemiology, clinical manifestations, and treatment options for skin and soft tissue infection caused by community-acquired methicillin-resistant Staphylococcus aureus. *J Am Acad Nurse Pract.* **2008,** *20* (2), 85-92.
11. Rutherford, S. T.; Bassler, B. L., Bacterial quorum sensing: its role in virulence and possibilities for its control. *Cold Spring Harb Perspect Med.* **2012,** *2* (11).
12. Sun, F.; Liang, H.; Kong, X.; Xie, S.; Cho, H.; Deng, X.; Ji, Q.; Zhang, H.; Alvarez, S.; Hicks, L. M.; Bae, T.; Luo, C.; Jiang, H.; He, C., Quorum-sensing agr mediates bacterial oxidation response via an intramolecular disulfide redox switch in the response regulator AgrA. *Proc Natl Acad Sci U S A.* **2012,** *109* (23), 9095-1000.
13. Kavanaugh, J. S.; Thoendel, M.; Horswill, A. R., A role for type I signal peptidase in Staphylococcus aureus quorum sensing. *Mol Microbiol.* **2007,** *65* (3), 780-798.

14. Malone, C. L.; Boles, B. R.; Horswill, A. R., Biosynthesis of Staphylococcus aureus autoinducing peptides by using the synechocystis DnaB mini-intein. *Appl Environ Microbiol.* **2007,** *73* (19), 6036-6044.
15. Yarwood, J. M.; Schlievert, P. M., Quorum sensing in Staphylococcus infections. *Journal of Clinical Investigation.* **2003,** *112* (11), 1620-1625.
16. Remy, B.; Mion, S.; Plener, L.; Elias, M.; Chabriere, E.; Daude, D., Interference in Bacterial Quorum Sensing: A Biopharmaceutical Perspective. *Front Pharmacol.* **2018,** *9*, 203.
17. Bassler, M. B. M. a. B. L., QUORUM SENSING IN BACTERIA. *Annu. Rev. Microbiol.* **2001,** *55:165–99*.
18. Figueroa, M.; Jarmusch, A. K.; Raja, H. A.; El-Elimat, T.; Kavanaugh, J. S.; Horswill, A. R.; Cooks, R. G.; Cech, N. B.; Oberlies, N. H., Polyhydroxyanthraquinones as quorum sensing inhibitors from the guttates of Penicillium restrictum and their analysis by desorption electrospray ionization mass spectrometry. *J Nat Prod.* **2014,** *77* (6), 1351-1358.
19. Daly, S. M.; Elmore, B. O.; Kavanaugh, J. S.; Triplett, K. D.; Figueroa, M.; Raja, H. A.; El-Elimat, T.; Crosby, H. A.; Femling, J. K.; Cech, N. B.; Horswill, A. R.; Oberlies, N. H.; Hall, P. R., omega-Hydroxyemodin limits staphylococcus aureus quorum sensing-mediated pathogenesis and inflammation. *Antimicrob Agents Chemother.* **2015,** *59* (4), 2223-2235.
20. Cech, N. B.; Junio, H. A.; Ackermann, L. W.; Kavanaugh, J. S.; Horswill, A. R., Quorum quenching and antimicrobial activity of goldenseal (Hydrastis canadensis) against methicillin-resistant Staphylococcus aureus (MRSA). *Planta Med.* **2012,** *78* (14), 1556-1561.
21. Cech, N. B.; Horswill, A. R., Small-molecule quorum quenchers to prevent Staphylococcus aureus infection. *Future Microbiol.* **2013,** *8* (12), 1511-1514.
22. Somerville, G. A.; Beres, S. B.; Fitzgerald, J. R.; DeLeo, F. R.; Cole, R. L.; Hoff, J. S.; Musser, J. M., In Vitro Serial Passage of Staphylococcus aureus: Changes in Physiology, Virulence Factor Production, and agr Nucleotide Sequence. *Journal of Bacteriology.* **2002,** *184* (5), 1430-1437.
23. Matthew Thoendel, J. S. K., Caralyn E. Flack, and Alexander R. Horswill*, Peptide Signaling in the Staphylococci. *Chem. Rev.* 2011, 111, 117–151.
24. Olson, M. E.; Todd, D. A.; Schaeffer, C. R.; Paharik, A. E.; Van Dyke, M. J.; Buttner, H.; Dunman, P. M.; Rohde, H.; Cech, N. B.; Fey, P. D.; Horswill, A. R., Staphylococcus epidermidis agr quorum-sensing system: signal identification, cross talk, and importance in colonization. *J Bacteriol.* **2014,** *196* (19), 3482-93.
25. Muhs, A.; Lyles, J. T.; Parlet, C. P.; Nelson, K.; Kavanaugh, J. S.; Horswill, A. R.; Quave, C. L., Virulence Inhibitors from Brazilian Peppertree Block Quorum Sensing and Abate Dermonecrosis in Skin Infection Models. *Sci Rep.* **2017,** *7*, 422-475.
26. Todd, D. A.; Parlet, C. P.; Crosby, H. A.; Malone, C. L.; Heilmann, K. P.; Horswill, A. R.; Cech, N. B., Signal Biosynthesis Inhibition with Ambuic Acid as

a Strategy To Target Antibiotic-Resistant Infections. *Antimicrob Agents Chemother.* **2017,** *61* (8).
27. Salam, A. M.; Quave, C. L., Targeting Virulence in Staphylococcus aureus by Chemical Inhibition of the Accessory Gene Regulator System In Vivo. *mSphere.* **2018,** *3* (1).
28. Quave, C. L.; Horswill, A. R., Flipping the switch: tools for detecting small molecule inhibitors of staphylococcal virulence. *Front Microbiol.* **2014,** *5*, 706.
29. Nakayama, J.; Uemura, Y.; Nishiguchi, K.; Yoshimura, N.; Igarashi, Y.; Sonomoto, K., Ambuic acid inhibits the biosynthesis of cyclic peptide quormones in gram-positive bacteria. *Antimicrob Agents Chemother.* **2009,** *53* (2), 580-6.
30. Xu, J.; Yang, X.; Lin, Q., Chemistry and biology of Pestalotiopsis-derived natural products. *Fungal Diversity.* **2014,** *66* (1), 37-68.
31. Strobel, G. A., Rainforest endophytes and bioactive products. *Crit Rev Biotechnol.* **2002,** *22* (4), 315-33.
32. J.Y. Li, J. K. H., David M. Grant, Bob Oka Tombe, Bharat Bashyal, W.M. Hess, Gary A. Strobel Ambuic acid , a highly functionalized cyclohexenone with antifungal activity from Pestalotiopsis spp. and Monochaeti. *Phytochem.* **2001,** (56), 463-468.
33. Kellogg, J. J.; Todd, D. A.; Egan, J. M.; Raja, H. A.; Oberlies, N. H.; Kvalheim, O. M.; Cech, N. B., Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds. *J. Nat. Prod.* **2016,** *79* (2), 376-86
34. Paguigan, N. D.; El-Elimat, T.; Kao, D.; Raja, H. A.; Pearce, C. J.; Oberlies, N. H., Enhanced dereplication of fungal cultures via use of mass defect filtering. *J Antibiot (Tokyo)* **2017,** *70* (5), 553-561.
35. Karen M VanderMolen, H. A. R., Tamam El-Elimat and Nicholas H Oberlies*, Evaluation of culture media for the production of secondary metabolites in a natural products screening program. *AMB Express.* **2013,** *3* (71).
36. El-Elimat, T.; Figueroa, M.; Ehrmann, B. M.; Cech, N. B.; Pearce, C. J.; Oberlies, N. H., High-resolution MS, MS/MS, and UV database of fungal secondary metabolites as a dereplication protocol for bioactive natural products. *J Nat Prod.* **2013,** *76* (9), 1709-16.
37. Hautbergue, T.; Jamin, E. L.; Debrauwer, L.; Puel, O.; Oswald, I. P., From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites. *Nat Prod Rep.* **2018,** *35* (2), 147-173.
38. Crusemann, M.; O'Neill, E. C.; Larson, C. B.; Melnik, A. V.; Floros, D. J.; da Silva, R. R.; Jensen, P. R.; Dorrestein, P. C.; Moore, B. S., Prioritizing Natural Product Diversity in a Collection of 146 Bacterial Strains Based on Growth and Extraction Protocols. *J Nat Prod.* **2017,** *80* (3), 588-597.
39. Caesar, L. K.; Kellogg, J. J.; Kvalheim, O. M.; Cech, R. A.; Cech, N. B., Integration of Biochemometrics and Molecular Networking to Identify Antimicrobials in Angelica keiskei. *Planta Med.* **2018,** *84* (9-10), 721-728.

40. Quinn, R. A.; Nothias, L. F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P. C., Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends Pharmacol Sci.* **2017,** *38* (2), 143-154.
41. Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W. T.; Crusemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderon, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C. C.; Floros, D. J.; Gavilan, R. G.; Kleigrewe, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C. C.; Yang, Y. L.; Humpf, H. U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; P, C. A. B.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodriguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P. M.; Phapale, P.; Nothias, L. F.; Alexandrov, T.; Litaudon, M.; Wolfender, J. L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D. T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Muller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. O.; Pogliano, K.; Linington, R. G.; Gutierrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N., Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* **2016,** *34* (8), 828-837.
42. Thurman, E. M.; Ferrer, I., The isotopic mass defect: a tool for limiting molecular formulas by accurate mass. *Anal Bioanal Chem.* **2010,** *397* (7), 2807-16.
43. Sleno, L., The use of mass defect in modern mass spectrometry. *J. Mass. Spectrom.* **2012,** 47, 226–236**,** *47*, 226–236.
44. Zhang, H.; Zhang, D.; Ray, K.; Zhu, M., Mass defect filter technique and its applications to drug metabolite identification by high-resolution mass spectrometry. *J Mass Spectrom.* **2009,** *44* (7), 999-1016.
45. Bertrand, S.; Bohni, N.; Schnee, S.; Schumpp, O.; Gindro, K.; Wolfender, J. L., Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery. *Biotechnol Adv.* **2014,** *32* (6), 1180-204.
46. Ekanayaka, E. A.; Celiz, M. D.; Jones, A. D., Relative mass defect filtering of mass spectra: a path to discovery of plant specialized metabolites. *Plant Physiol.* **2015,** *167* (4), 1221-32.
47. M. Zhu, L. M., H. Zhang, W.G. Humphreys. , Detection and Structural Characterization of Glutathione-Trapped Reactive Metabolites Using Liquid

Chromatography-High-Resolution Mass Spectrometry and Mass Defect Filtering. *Anal. Chem.* **2007,** *79*, 8333.
48. da Silva, R. R.; Vargas, F.; Ernst, M.; Nguyen, N. H.; Bolleddu, S.; Del Rosario, K. K.; Tsunoda, S. M.; Dorrestein, P. C.; Jarmusch, A. K., Computational Removal of Undesired Mass Spectral Features Possessing Repeat Units via a Kendrick Mass Filter. *J Am Soc Mass Spectrom.* **2018**.
49. da Silva, R. R.; Dorrestein, P. C.; Quinn, R. A., Illuminating the dark matter in metabolomics. *Proc Natl Acad Sci U S A.* **2015,** *112* (41), 12549-50.
50. Bouslimani, A.; Sanchez, L. M.; Garg, N.; Dorrestein, P. C., Mass spectrometry of natural products: current, emerging and future technologies. *Nat Prod Rep.* **2014,** *31* (6), 718-29.
51. Kersten, R. D.; Yang, Y. L.; Xu, Y.; Cimermancic, P.; Nam, S. J.; Fenical, W.; Fischbach, M. A.; Moore, B. S.; Dorrestein, P. C., A mass spectrometry-guided genome mining approach for natural product peptidogenomics. *Nat Chem Biol.* **2011,** *7* (11), 794-802.
52. Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C., Molecular networking as a dereplication strategy. *J Nat Prod.* **2013,** *76* (9), 1686-1699.
53. Guthals, A.; Watrous, J. D.; Dorrestein, P. C.; Bandeira, N., The spectral networks paradigm in high throughput mass spectrometry. *Mol Biosyst.* **2012,** *8* (10), 2535-44.
54. Medema, M. H.; Fischbach, M. A., Computational approaches to natural product discovery. *Nat Chem Biol.* **2015,** *11* (9), 639-648.
55. Zhang, H.; Zhang, D.; Ray, K., A software filter to remove interference ions from drug metabolites in accurate mass liquid chromatography/mass spectrometric analyses. *J Mass Spectrom.* **2003,** *38* (10), 1110-1112.
56. Amrine, C. S. M.; Raja, H. A.; Darveaux, B. A.; Pearce, C. J.; Oberlies, N. H., Media studies to enhance the production of verticillins facilitated by in situ chemical analysis. *J Ind Microbiol Biotechnol.* **2018**.
57. Holman, J. D.; Tabb, D. L.; Mallick, P., Employing ProteoWizard to Convert Raw Mass Spectrometry Data. *Curr Protoc Bioinformatics.* **2014,** *46*, 13 24 1-9.
58. Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M., MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics.* **2010,** *11* (1), 395.
59. Pan, H., A non-covalent dimer formed in electrospray ionization mass spectrometry behaving as a precursor for fragmentations. *Rapid Commun Mass Spectrom.* **2008,** *22* (22), 3555-60